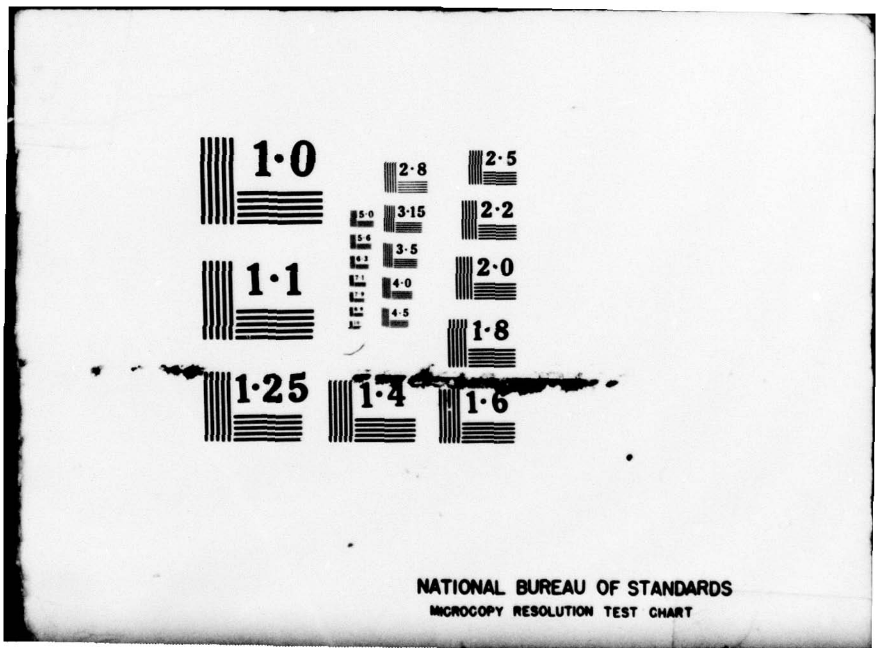
AD-A054 251 UNIVERSITY OF SOUTHERN CALIFORNIA LOS ANGELES DEPT OF--ETC F/G 12/1
FITTING A THEORETICAL DISTRIBUTION TO SUBJECTIVELY ASSESSED FRA--ETC(U)
1978 C LIN, G J SCHICK N00014-75-C-0733

UNCLASSIFIED NL

1 OF 1
ADA
054251

END
DATE
FILMED
6-78
DDC


1·0
2·8
2·5
3·15
2·2
3·5
2·0
1·1
4·0
4·5
1·8
1·25
1·4
1·6
NATIONAL BUREAU OF STANDARDS
MICROCOPY RESOLUTION TEST CHART

# FITTING A THEORETICAL DISTRIBUTION TO SUBJECTIVELY ASSESSED FRACTILES*

Chi-Yuan Lin and George J. Schick
University of Southern California

Abstract

A prior distribution is an indispensable input to Bayesian analysis. When a prior distribution is elicited from a decision maker (or his designated expert) based on his subjective judgment, it is not readily expressed by a mathematical function. To make Bayesian analysis tractable mathematically, it is essential to express the prior distribution by a mathematical function. This paper uses the lognormal distribution to investigate various methods of fitting a theoretical distribution to a subjective prior.

## 1. INTRODUCTION

A significant contribution of Bayesian decision theory is the formal procedure for considering the decision maker's a priori knowledge in a given situation. By the use of Bayes' Theorem, the decision maker's a priori knowledge (expressed in terms of a prior distribution) and the sample information (expressed in terms of a likelihood function) are combined to form a posterior distribution on which the decision is based.

In the past, most Bayesian statisticians concentrated on the mathematical derivation of the posterior distribution from a given prior and a given likelihood function. In many practical situations, however, the prior distribution has to be derived from the decision maker (or his designated expert) or from previous sample data. Thus, during recent years, researchers [3, 4, 5, 6, 7, 8] have turned their attention to assessments of prior distributions.

*This research was supported by the office of Naval Research under Contract N0014-75-C-0733, Task Nr. 042-323. Reproduction in whole or part is permitted for any purpose of the United States Government.

The prior distribution obtained subjectively from a decision maker is not expressed in mathematical form. In fact, the responses from a decision maker provide several fractile points of the distribution, rather than the entire distribution. However, in order to use Bayes' theorem to derive the posterior distribution, the prior distribution must be completely specified. Furthermore, to derive the posterior distribution mathematically, it is necessary to have the prior distribution expressed by a mathematical function. Therefore, the primary purpose of the paper is to examine alternative ways of estimating parameters of a probability distribution from subjectively assessed fractiles. Since the study was part of a research project in the field of reliability and maintainability, the lognormal distribution, which has been used extensively in this field, was selected to illustrate the procedure.

## 2. COMMONLY USED METHODS

The method of equally likely subintervals perhaps is the most commonly used approach in eliciting a subjective prior distribu-

tion. As explained by Lin and Schick [3], the basic idea of this method is to ask the decision maker, at any stage, to divide a given interval of a distribution into two judgmentally equally likely subintervals. Technically, if $x_k$ is used to designate the $k^{th}$ fractile of the probability distribution of an uncertain quantity (i.e., a random variable) $\tilde{x}$, then the decision maker is asked to respond to a series of questions from which the fractiles $x_{.5}$, $x_{.25}$, $x_{.75}$, . . .are successively ascertained.

Applying this method to determine the subjective distribution of the maintenance time (in minutes) for a particular job, five fractiles are obtained, as shown in Table 1. Also, the mode is determined subjectively as 275. With the fractiles and

Table 1
Subjective Fractiles

| Cumulative Probability | Corresponding x-value |
|---|---|
| .00 | 240 |
| .25 | 265 |
| .50 | 280 |
| .75 | 300 |
| 1.00 | 340 |

mode from the expert, the program developed by Lin and Schick [3] fits a smooth curve by successive use of quadratic functions. Then, by dividing this curve into 100 equally likely subintervals and regarding the "median" in each of these intervals as the representative value of that subinterval, the mean and standard deviation are computed. For this example the mean is 283.04 minutes and the standard deviation is 24.07 minutes.

One way of fitting a lognormal distribution to the subjective distribution obtained above is to match the mean and standard deviation of the lognormal distribution with the mean and standard deviation of the subjective distribution. To do this, first recall that [1] if the lognormal density function with parameters $\alpha$ and $\beta$ is written as:

$$f(x|\alpha,\beta) = \frac{1}{\beta\sqrt{2\pi}}x^{-1}\exp\left[-\tfrac{1}{2}\left(\frac{\ln x-\alpha}{\beta}\right)^2\right], \begin{matrix}x>0\\ \beta>0\end{matrix} \quad (1)$$

then its mean and standard deviation are:

$$\mu = \exp(\alpha + \tfrac{1}{2}\beta^2) \quad (2)$$

$$\sigma = \mu(\exp \beta^2 - 1)^{\frac{1}{2}} \quad (3)$$

Solving these two equations simultaneously for $\alpha$ and $\beta$, we obtain:

$$\alpha = \ln \mu - \tfrac{1}{2}\beta^2 \quad (4)$$

$$\beta = [\ln(1 + \sigma^2/\mu^2)]^{\frac{1}{2}} \quad (5)$$

Finally, substituting the mean and standard deviation of the subjective distribution ($\mu$ = 283.04 and $\sigma$ = 24.07) into equations (4) and (5), we obtain $\alpha$ = 5.6424 and $\beta$ = 0.0800. Therefore, the subjective distribution is fitted by the lognormal distribution with parameters $\alpha$ and $\beta$ thus obtained.

Since the lognormal distribution given by equation (1) has two-parameters, the fit can also be accomplished by matching any two selected values of the two distributions, such as .25 fractile and the mode, .90 fractile and the mean. Different selections will yield different values of the parameters $\alpha$ and $\beta$. Thus, an important question arises: "Is there a particular set of input values that will consistently yield the best fit?"

To investigate this question, twenty-eight sets of input values from each of two subjective distributions were used to determine the parameters of lognormal distributions. The relevant fractiles and parameters of the subjective distributions are presented in Table 2. The means of these distributions are almost equal, but distribution A is more homogeneous than distribution B.

Table 2
Selected Fractiles and Parameters of
Two Subjective Distributions

| Fractiles & Parameters | Subjective Dist. A | Subjective Dist. B |
|---|---|---|
| $x_0$ | 240 | 50 |
| $x_{.10}$ | 252.45 | 98.7 |
| $x_{.125}$ | 255 | 110 |
| $x_{.250}$ | 265 | 160 |
| $x_{.375}$ | 272 | 200 |
| $x_{.50}$ | 280 | 250 |
| $x_{.625}$ | 289 | 320 |
| $x_{.75}$ | 300 | 400 |
| $x_{.875}$ | 315 | 500 |
| $x_{.90}$ | 319.1 | 525.5 |
| $x_{1.00}$ | 340 | 650 |
| Mode | 275 | 200 |
| $\mu$ | 283.0439 | 284.9094 |
| $\sigma$ | 24.0676 | 155.4427 |

One of the two statistical measures chosen to determine the goodness of fit is the Kolmogorov-Smirnov (K - S) test statistic:

$$D = \max_{\forall i} |o_i - e_i| \tag{6}$$

where $o_i$ are the observed values and $e_i$ are expected values. The results indicate that no particular pair of input values will consistently yield the best fit. Notice that the use of K - S statistic here is not a usual application of goodness-of-fit test. We simply use K - S statistics for purposes of comparison rather than for statistical tests.

The other measure used to compare the goodness of fit is the Average Absolute Deviation:

$$AD = \frac{1}{n} \sum_{i=1}^{n} |o_i - e_i| \tag{7}$$

The results reveal that no particular pair of input values consistently yields the best fit.

Since test results did not detect a particular pair of input values that would consistently yield the best fit, an important practical question arises: "Which pair of values should be used in fitting a theoretical distribution (such as a lognormal) to a subjective distribution?" Conceivably, one could use various pairs of values to obtain different distributions and choose the one that would yield the best fit according to a prespecified criterion such as K - S statistics or average absolute deviation. However, this would be time-consuming and costly. Thus, an alternative fitting procedure is desirable.

## 3. PROPOSED METHOD

Although only two values of a subjective distribution are needed for estimating the parameters of a two-parameter lognormal distribution, more than two values may be used. It would seem reasonable to assume that a method which makes use of more than two points to fit a distribution would yield a consistently better fit than another method that uses only two points. Thus, rather than fitting a lognormal distribution to two selected values of the subjective distribution that has been obtained by smoothing the subjective fractile points, we can use the least squares method to fit a lognormal distribution directly to the subjective fractiles.

By using the least-squares method to fit a lognormal distribution to the 0, .125, .25, .375, .50, .625, .75, .875 and 1.0 subjective fractiles of Distribution A, we obtain the parameters of the lognormal distribution as: $\alpha$ = 5.6388, $\beta$ = .0928. Following the same procedure, we obtain the lognormal parameters for Distribution B as: $\alpha$ = 5.5515, $\beta$ = .6454. Compared with the previous method, the Least-Squares estimates of $\alpha$ and $\beta$ for Distribution A yield an average deviation that is smaller than any of the 28 average deviations obtained previously. Although in two of the 25 cases for Distribution B, the average deviations obtained by the previous method

are smaller than the average deviation obtained by the Least-Squares method, the differences are negligible for practical purposes. The K - S statistic obtained for the Least-Squares method is smaller than those obtained previously in 18 out of 28 cases for Distribution A and in 20 out of 25 cases for Distribution B. However, in those cases in which the K - S statistic for Least-Squares method is larger, the differences are rather small. Furthermore, many of the K - S statistics turn out to be the deviations at the tails of distributions.

## 4. CONCLUSION

To fit a two-parameter lognormal distribution to a subjective distribution, two selected values of the subjective distribution, such as mean and standard deviation, or the .25 fractile and mode will be sufficient to determine the two parameters and hence to completely specify the lognormal distribution. The study has revealed that different sets of input values yield different goodness of fit. Furthermore, using either the Kolmogorov-Smirnov statistic or the average absolute deviation as a measure of goodness of fit, the study has also found that no particular set of values consistently yields the best fit. Thus, in applying this method to fit a lognormal distribution to a subjective prior, it is difficult to decide which pair of values to use. To resolve this difficulty, the paper has presented an alternative procedure which uses the least-squares method to fit a lognormal distribution directly to the subjective fractiles rather than the two points of the subjective distribution that has been obtained by smoothing the fractiles. This method yields a better fit than the previous method which requires the selection of pairs of input values.

## REFERENCES


1. Aitchison, J. and J. A. C. Brown, The Lognormal Distribution, London: Cambridge University Press, 1957.

2. Howard, Ronald A., "The Foundations of Decision Analysis," IEEE Transactions on Systems Science and Cybernetics, Vol. SSC-4, No. 3, September 1968.

3. Lin, Chi-Yuan, and Schick, George J., "Determination of Subjective Prior Distribution for Bayesian Analysis of Maintainability Problems," Annals of Reliability and Maintainability, Vol. 13, 1974.

4. Pratt, J. W., Raiffa, H., and Schlaifer, R., Introduction to Statistical Decision Theory (preliminary edition), McGraw-Hill, New York, 1965.

5. Raiffa, Howard, Decision Analysis: Introductory Lectures on Choices under Uncertainty, Addison-Wesley, Reading, Massachusetts, 1968.

6. Savage, Leonard J., "The Elicitation of Personal Probabilities and Expectations," Journal of the American Statistical Association, Vol. 66, pp.783-801, 1971.

7. Schlaifer, Robert, Analysis of Decisions Under Uncertainty, McGraw-Hill, New York, 1969.

8. Winkler, Robert L., "The Assessment of Prior Distributions in Bayesian Analysis," Journal of the American Statistical Association, Vol. 62, pp.776-800, 1967.


## BIOGRAPHY


Chi-Yuan Lin received his Ph.D. from MIT. He is an Associate Professor of Decision Sciences and Director of the Doctoral Program in the Graduate School of Business Administration at USC. His primary research interest is in the fields of decision analysis and statistics.

George J. Schick received his Ph.D. from the University of Wisconsin and is a Professor of Decision Sciences at USC. His research interest is in the fields of reliability and decision analysis.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

# REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

| | |
|---|---|
| 1. REPORT NUMBER | 2. GOVT ACCESSION NO. |
| 3. RECIPIENT'S CATALOG NUMBER | 9 |
| 4. TITLE (and Subtitle) | 6 FITTING A THEORETICAL DISTRIBUTION TO SUBJECTIVELY ASSESSED FRACTILES. |
| 5. TYPE OF REPORT & PERIOD COVERED | Technical rept., |
| 6. PERFORMING ORG. REPORT NUMBER | |
| 7. AUTHOR(s) | 10 Chi-Yuan Lin<br>George J. Schick |
| 8. CONTRACT OR GRANT NUMBER(s) | 15 N00014-75-C-0733 new |
| 9. PERFORMING ORGANIZATION NAME AND ADDRESS | Decision Sciences ~~Department~~ Dept of<br>University of Southern California<br>Los Angeles, CA 90007 |
| 10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS | NR042-323 |
| 11. CONTROLLING OFFICE NAME AND ADDRESS | Office of Naval Research<br>Statistics and Probability Program, Code 436<br>Arlington, Virginia 22217 |
| 12. REPORT DATE | 11 1978 |
| 13. NUMBER OF PAGES | 4 |
| 14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office) | 12 5p. |
| 15. SECURITY CLASS. (of this report) | Unclassified |
| 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE | |

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Bayesian statistical analysis
Assessment of prior distributions
Lognormal distribution

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

A prior distribution is an indispensable input to Bayesian analysis. When a prior distribution is elicited from a decision amker (or his designated expert) based on his subjective judgement, it is not readily expressed by a mathematical function. To make Bayesian analysis tractable mathematically, it is essential to express the prior distribution by a mathematical function. This paper uses the lognormal distribution to investigate various methods of fitting a theoretical distribution to a subjective prior.



New 410692